DISTRIBUTED COMPRESSION OF THE PLENOPTIC FUNCTION

Nicolas Gehrig and Pier Luigi Dragotti

Communications and Signal Processing Group, Electrical and Electronic Engineering Department
Imperial College, Exhibition Road, London SW7 2AZ, United Kingdom
e-mail: {nicolas.gehrig, p.dragotti}@imperial.ac.uk

ABSTRACT

In this paper, we consider the problem of distributed compression in camera sensor networks. Due to the spatial proximity of the different cameras, acquired images can be highly dependent. The correlation in the visual information retrieved is related to the structure of the plenoptic function and can be estimated using geometrical information such as the position of the cameras and some bounds on the location of the objects.

We propose a distributed compression scheme that takes advantage of this geometrical information in order to reduce the overall transmission rate from the sensors to a common central receiver. This new approach allows for a flexible repartition of the transmission bit-rates amongst the encoders and is optimal in many cases. Moreover, we show that our coding scheme can be made resilient to a fixed number of occlusions and that perfect reconstruction and interpolation are possible at the receiver.

1. INTRODUCTION

The recent emergence of sensor network technology has an inevitable impact on the need for new distributed approaches in most areas of signal processing. Problems such as sampling, approximation, compression and reconstruction have been extensively studied in the traditional centralized context, but theories resulting from this research are usually not directly applicable to distributed schemes. The idea of sensor network is to replace the common centralized scenario by a new completely distributed approach for acquiring and processing data. It consists of numbers of independent sensors, densely deployed, having processing and communication capabilities. Our work focuses on camera sensor networks, where each sensor is a self-powered wireless device containing a digital camera and a processing unit. We assume that our sensors are looking at a specific scene from different view-points and transmit their information to a common central receiver.

Due to their spatial proximity, images acquired by two different sensors can be highly dependent. Since the inter-sensor communication can be extremely expensive in terms of power consumption, we would like to achieve the best possible transmission rate (from the sensors to the common receiver) by exploiting the overall correlation without allowing the sensors to communicate with each other. Thanks to results obtained by Slepian-Wolf [1] and Wyner-Ziv [2], we know that in many cases, we can theoretically achieve the compression rate of a joint-encoder using separate encoders. However, this surprising result assumes that the correlation structure of the sources is a priori known at each individual encoder. Although the theoretical aspect of distributed source coding has been known for about three decades, it is only recently that practical coding approaches have been proposed. In [3], Pradhan and Ramchandran proposed a coding technique inspired from channel coding. Practical designs mainly based on turbo and LDPC codes have since been presented in several other papers (see [4, 5, 6] for example).

In this paper, we consider a simplified geometrical set-up and show that the correlation structure of the source, which is given by the plenoptic function [7], can be precisely estimated using some a priori global geometrical information. For instance, the location of the cameras might be known and some objects of interest might be well localized in space. We then propose a distributed compression algorithm that exploits this a priori knowledge. This new scheme allows for a flexible repartition of the bit-rates amongst the encoders and is optimal in many cases (i.e., it achieves the Slepian and Wolf performance (see Proposition 1)). We then show that our coding scheme can be made resilient to a fixed number of occlusions (see Proposition 2) and that perfect reconstruction and interpolation are possible at the receiver.

The paper is organized as follows: The next section introduces the plenoptic function and gives a precise description of our problem statement. In Section 3, we present our distributed compression approach and we address the problem of occlusions. Simulation results can be found in Section 4 and we conclude in Section 5.

2. THE PLENOPTIC FUNCTION AND OUR CAMERA SENSOR NETWORK

The plenoptic function was first introduced by Adelson and Bergen in 1991 [7]. It corresponds to the function representing the intensity and chromaticity of the light observed from every position and direction in the 3D space, and can therefore be parameterized as a 7D function: $P_7 = P(\theta, \phi, \lambda, t, V_x, V_y, V_z)$.

A camera sensor network is able to acquire a finite number of different views of a scene at any given time and can thus be seen as a sampling device for the plenoptic function. We choose the following scenario for our work: Assume that we have N cameras evenly placed on a horizontal line. Let α be the distance between two consecutive cameras, and assume that they are all looking in the same direction (perpendicular to the line of cameras). Assume then that the observed scene is composed of simple objects such as uniformly colored polygons parallel to the image plane and with depths bound between the two values z_{min} and z_{max} as shown in Figure 1. According to the epipolar geometry principles, which are directly related to the structure of the plenoptic function (see Figure 2), we know that the difference between the positions of a specific object on the images obtained from two consecutive cam-

This work was supported in part by DIF-DTC project number 12.6.2. and the Gabor Fellowship (EEE Department).

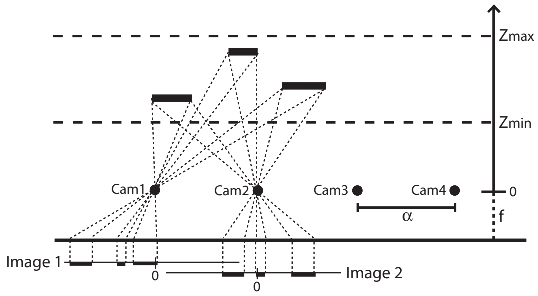

Fig. 1. Our camera sensor network configuration.

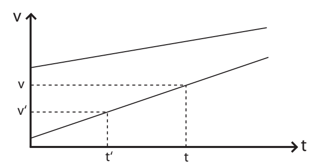

Fig. 2. 2D plenoptic function of two points. The t-axis corresponds to the camera position and v corresponds to the relative positions on the corresponding image. A point of the scene is therefore represented by a line whose slope is directly related to the point's depth. The difference between the positions of a given point on two different images thus satisfies the relation $(v - v') = \frac{f(t-t')}{z}$, where z is the point's depth and f is the focal length of the cameras.

eras will be equal to $\Delta = \frac{\alpha f}{z}$, where z is the depth of the object and f is the focal length of the cameras. This disparity Δ depends only on the distance z of the point from the focal plane. If we know a priori that there is a finite depth of field, that is $z \in [z_{min}, z_{max}]$, then there is a finite range of disparities to be coded, irrespective of how complicated the scene is. This key insight is used in this paper to develop new distributed compression algorithms as we show in the next section.

Notice that a similar insight has been previously used by Chai et al. to develop new schemes to sample the plenoptic function [8].

3. DISTRIBUTED COMPRESSION

3.1. Background

Consider a communication system where two discrete correlated sources X and Y are to be encoded at rates R_X and R_Y respectively, and transmitted to a central receiver. If it were possible to perform the coding jointly, a rate $R_X + R_Y \geq H(X,Y)$ would be sufficient to perform noiseless coding. Now assume that these two sources are physically separated and cannot communicate with one another, Slepian and Wolf [1] showed that lossless compression of X and Y is still achievable if $R_X \geq H(X|Y)$, $R_Y \geq H(Y|X)$ and $R_X + R_Y \geq H(X,Y)$. This means that there is no loss in terms of overall rate even though the encoders are separated (see Figure 3).

In the next subsection, we propose a distributed coding scheme for the configuration presented in Figure 1 with two cameras. Since both encoders have some knowledge about the geometry of the

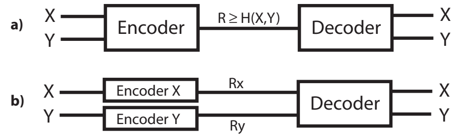

Fig. 3. (a) Joint source coding. (b) Distributed source coding. The Slepian-Wolf theorem (1973) states that a combined rate of $H(X,Y)$ remains sufficient even if the correlated signals are encoded separately. The achievable rate region is given by: $R_X \geq H(X|Y)$, $R_Y \geq H(Y|X)$ and $R_X + R_Y \geq H(X,Y)$.

scene, the correlation structure of the two sources can be easily retrieved. We then show that our coding technique can be used with any pair of bit-rates contained in the achievable rate region defined by Slepian and Wolf, and can therefore be optimal.

3.2. Asymmetric and Symmetric encoding with two cameras

Let X and Y be the horizontal positions of a specific object on the images obtained from two consecutive cameras. Assume the image width is made of 2^R pixels. Due to the epipolar geometry and the information we have about the scene, that is $(\alpha, f, z_{min}, z_{max})$, we know that $Y \in [X + \frac{\alpha f}{z_{max}}, X + \frac{\alpha f}{z_{min}}]$ for a specific X. Encoding X and Y independently would require a total of $H(X) + H(Y)$ bits. However, using a *coset* approach, we can modulo encode Y as $Y' = Y \mathrm{mod} \lceil \alpha f(\frac{1}{z_{min}} - \frac{1}{z_{max}}) \rceil$. The receiver will then retrieve the correct Y corresponding to the received Y' such that $Y \in [X + \frac{\alpha f}{z_{max}}, X + \frac{\alpha f}{z_{min}}]$. The overall transmission rate is therefore decreased to $H(X) + H(Y')$ bits. If we assume that the difference between X and Y is uniformly distributed in $[\frac{\alpha f}{z_{max}}, \frac{\alpha f}{z_{min}}]$, we can claim that $H(Y') = H(Y|X)$. We can therefore see that our coding scheme using $H(X) + H(Y') = H(X) + H(Y|X) = H(X,Y)$ bits is optimal.

This simple distributed coding technique is very powerful since it takes full advantage of the geometrical information to minimize the global transmission bit-rate. However, its asymmetric repartition of the bit-rates may be problematic for some practical applications. In the following, we will show that our coding approach can be extended in a way such that any pair of bit-rates satisfying the Slepian and Wolf conditions can be used. Looking at the following relation: $H(X,Y) = H(X|Y) + H(Y|X) + I(X,Y)$, we can see that the minimum information that must be sent from the source X corresponds to the conditional entropy $H(X|Y)$. Similarly, the information corresponding to $H(Y|X)$ must be sent from the source Y. The remaining information required at the receiver in order to recover the values of X and Y perfectly is related to the mutual information $I(X,Y)$ and is by definition available at both sources. This information can therefore be obtained partially from both sources in order to balance the transmission rates.

We know that the correlation structure between the two sources is such that Y belongs to $[X + \frac{\alpha f}{z_{max}}, X + \frac{\alpha f}{z_{min}}]$ for a given X. Let $\widetilde{Y}$ be defined as $\widetilde{Y} = Y - \lceil \frac{\alpha f}{z_{max}} \rceil$. This implies that the difference $(\widetilde{Y} - X)$ is contained in $\{0, 1, \ldots, \delta\}$, where $\delta = \lceil \alpha f(\frac{1}{z_{min}} - \frac{1}{z_{max}}) \rceil$. Looking at the binary representations of X and $\widetilde{Y}$, we can say that the difference between them can be computed using only their last R_{min} bits where $R_{min} = \lceil \log_2(\delta + 1) \rceil$. Let X_1

and $\widetilde{Y}_1$ correspond to the last R_{min} bits of X and $\widetilde{Y}$ respectively. Let $X_2 = (X \gg R_{min})$ and $\widetilde{Y}_2 = (\widetilde{Y} \gg R_{min})$, where the "$\gg$" operator corresponds to a binary shift to the right. We can thus say that $\widetilde{Y}_2 = X_2$ if $\widetilde{Y}_1 \geq X_1$ and that $\widetilde{Y}_2 = X_2 + 1$ if $\widetilde{Y}_1 < X_1$. As presented in Figure 4, our coding strategy consists in sending X_1 and $\widetilde{Y}_1$ from the sources X and Y respectively and then, sending only a subset of the bits for X_2 and only the complementary one for $\widetilde{Y}_2$. At the receiver, X_1 and $\widetilde{Y}_1$ are then compared to determine if $\widetilde{Y}_2 = X_2$ or if $\widetilde{Y}_2 = X_2 + 1$. Knowing this relation and their partial binary representations, the decoder can now perfectly recover the values of X and $\widetilde{Y}$.

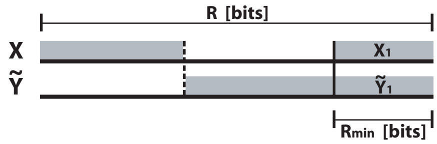

Fig. 4. Binary representation of the two correlated sources. The last R_{min} bits are sent from the two sources but only complementary subsets of the first $(R - R_{min})$ bits are necessary at the receiver for a perfect reconstruction of X and Y.

Assume that z_{min} and z_{max} are such that $(\delta + 1)$ is a power of 2. Since we assume that $(\widetilde{Y} - X)$ is uniformly distributed, we can state that $H(\widetilde{Y} - X) = H(X|Y) = H(Y|X) = R_{min}$. Let $S(X_2)$ be a subset of the $R - R_{min}$ bits of X_2 and let $\bar{S}(\widetilde{Y}_2)$ corresponds to the complementary subset of $\widetilde{Y}_2$. If we assume now that X is uniformly distributed in $\{0, 1, \ldots, 2^R - 1\}$, we can say that $H(S(X_2)) + H(\bar{S}(\widetilde{Y}_2)) = H(S(X_2), \bar{S}(\widetilde{Y}_2)) = I(X, Y)$. The total rate necessary for our scheme corresponds to $I(X, Y) + 2R_{min} = H(X, Y)$ and is therefore optimal. We can now summarize our results into the following proposition:

Proposition 1 *Consider the configuration presented in Figure 1 with two cameras, and assume that no occlusion happens in the two corresponding views. The following distributed coding strategy is sufficient to allow for a perfect reconstruction of these two views at the decoder. For each object's position:*

- *Send the last R_{min} bits from both sources, with $R_{min} = \lceil \log_2(\delta + 1) \rceil$ and $\delta = \lceil \alpha f(\frac{1}{z_{min}} - \frac{1}{z_{max}}) \rceil$.*
- *Send complementary subsets for the first $(R - R_{min})$ bits.*

If we assume that X and $(Y - X)$ are uniformly distributed and that $\delta = 2^{R_{min}} - 1$, this coding strategy achieves the Slepian-Wolf bounds and is therefore optimal.

3.3. Three cameras with one possible occlusion

In order to reconstruct the position of an object for any virtual camera position, we need to know its correct position in at least two different views. Using the epipolar geometry principles, we can then easily retrieve its absolute position and depth. Unfortunately, a specific object may not be visible from certain view points since it might be hidden behind another object or might be out of field. Nevertheless, using a configuration with more cameras will make it more likely for any object to be visible in at least two views.

Assume we have three cameras in a configuration similar to the one presented in Figure 1 and that each object of the scene can be occluded in at most one of these three views. Our goal is to design a distributed coding scheme for these three correlated sources such that the information provided by any pair of these sources is sufficient to allow for a perfect reconstruction at the receiver. Let X, Y and Z be the horizontal positions of a specific object on the images obtained from camera 1, 2 and 3 respectively. We know that Y belongs to $[X + \frac{\alpha f}{z_{max}}, X + \frac{\alpha f}{z_{min}}]$ and Z belongs to $[X + 2\frac{\alpha f}{z_{max}}, X + 2\frac{\alpha f}{z_{min}}]$ for a given X. Moreover, we know that any of these variables is deterministic given the two others and follows the relation $Z = 2Y - X$. Let $\widetilde{X}$ and $\widetilde{Z}$ be defined as $\widetilde{X} = X + \frac{\alpha f}{z_{mean}}$ and $\widetilde{Z} = Z - \frac{\alpha f}{z_{mean}}$ where z_{mean} is defined such that $\frac{1}{z_{mean}} = \frac{1}{2}(\frac{1}{z_{min}} + \frac{1}{z_{max}})$. It implies that the differences $(Y - \widetilde{X})$ and $(\widetilde{Z} - Y)$ are equal and belong to $[-\delta/2, \delta/2]$ and that the difference $(\widetilde{Z} - \widetilde{X})$ belongs thus to $[-\delta, \delta]$, where δ is defined as in Section 3.2.

Looking at the binary representation of $\widetilde{X}$, Y and $\widetilde{Z}$ (at integer precision), we can say that the difference between any pair can be retrieved using only their last R_{min} bits, where $R_{min} = \lceil \log_2(2\delta + 1) \rceil$. Let $\widetilde{X}_1$, Y_1 and $\widetilde{Z}_1$ correspond to the last R_{min} bits of $\widetilde{X}$, Y and $\widetilde{Z}$ respectively. Using a similar approach to that presented in Section 3.2, we know that any complementary binary subsets of $\widetilde{X}_2$, Y_2 and $\widetilde{Z}_2$ are necessary at the receiver to allow for a perfect reconstruction. Since one occlusion can happen, we need to choose the binary subsets such that any pair of these subsets contains at least one value for each of the $(R - R_{min})$ bits. A possible repartition is shown in Figure 5 (symmetric case). A transmission rate of $\frac{2}{3}r + R_{min}$ for each source is necessary in this case, where $r = R - R_{min}$.

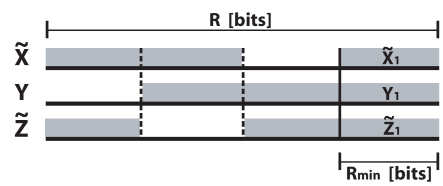

Fig. 5. Binary representation of the three correlated sources. The last R_{min} bits are sent from the three sources but only subsets of the first $(R - R_{min})$ bits are necessary at the receiver for a perfect reconstruction of X, Y and Z even if one occlusion occurs.

On receiving the last R_{min} bits from only two sources, the decoder is able to retrieve the last R_{min} bits of the third one, which may be occluded. Therefore, the relationship between $\widetilde{X}_2$, Y_2 and $\widetilde{Z}_2$ can be obtained and only subsets of their binary representations are necessary for a perfect reconstruction. Since an occlusion may have occurred, each bit position has been sent from two different sources, which implies a global transmission of $2r$ bits for the first $(R - R_{min})$ bits. It is therefore apparent that our total transmission rate of $\frac{2}{3}r + R_{min}$ bits per sources was therefore optimal.

We can now generalize our result (see Figure 6) to any number of cameras and occlusions with the following proposition:

Proposition 2 *Consider a system with N cameras as depicted in Figure 1. Assume that any object of the scene can be occluded in at most $M \leq N - 2$ views. The following distributed coding strategy is sufficient to allow for a perfect reconstruction of these N views at the decoder and to interpolate any new view:*

- *Send the last R_{min} bits of the objects'positions from only the first $(M+2)$ sources, with $R_{min} = \lceil \log_2((M+1)\delta) \rceil$ and $\delta = \lceil \alpha f(\frac{1}{z_{min}} - \frac{1}{z_{max}}) \rceil$.*
- *For each of the N sources, send only a subset of its first $(R - R_{min})$ bits such that each particular bit position is sent from exactly $(M+1)$ sources.*

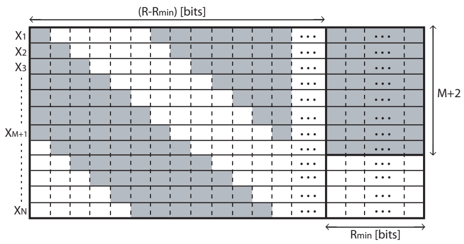

Fig. 6. Binary representation of the N correlated sources. The last R_{min} bits are sent only from the $(M+2)$ first sources. Only subsets of the first $(R - R_{min})$ bits are sent from each source, such that each bit position is sent exactly from $(M+1)$ sources.

Notice that our distributed coding strategy can be generalized to any binary sources. We proposed a general extension based on linear channel codes and syndrome encoding [9, 10].

4. SIMULATION RESULTS

We developed a simulation to illustrate the performance of our distributed compression scheme. We created an artificial scene composed of simple objects such as polygons of different intensities placed at different depths. Our system could then generate any view of that scene for any specified camera position. In the example presented in Figure 7, we generated three views of a simple scene composed of three objects such that one of them is occluded in the second view, and another one is out of field in the third view. The three generated images have a resolution of 512×512 pixels and are used as the inputs for the testing of our distributed compression algorithm. Each encoder applies first a simple corner detection to retrieve the vertex positions of their visible polygons. Each vertex (x, y) is represented using $2R = 2\log_2(512) = 18$ bits. Each encoder knows the relative locations of the two other cameras ($\alpha = 100$) but does not know the location of the objects. It only knows that the depths of the objects are contained in $[1.95, 5.05]$ and that $f = 1$. Depending on its depth, an object will thus move from 20 to 51 pixels between two consecutive views. This means that the difference Δ on two consecutive positions can be described using $R_{min} = \log_2(51 - 19) = 5$ bits.

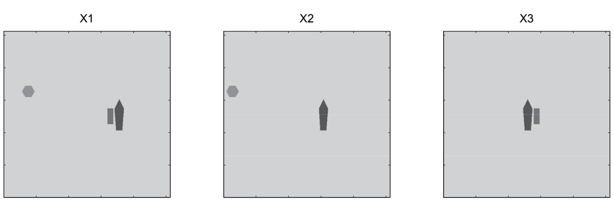

Fig. 7. Three views of a simple synthetic scene obtained from three aligned and evenly spaced cameras. Note that an occlusion happens in X_2 and that an object is out of field in X_3.

In order to be resilient to one occlusion, we applied the approach proposed in Section 3.3. The results showed that only 14 bits per vertex were necessary from each source (instead of 18) to allow for a perfect reconstruction of the scene at the receiver. When repeating the operation with three other views and assuming that no occlusion was possible, only 8 bits per vertex were necessary from each source.

5. CONCLUSIONS

We have proposed a new distributed compression scheme for camera sensor networks. Our method uses some geometrical information about the scene in order to estimate the plenoptic constraints and retrieve a correlation structure for the sources. A solution to the problem of occlusions has also been proposed. Ongoing research is focusing on the extension to more general scenes, leading to the development of efficient distributed compression algorithms for camera sensor networks in natural scenes.

6. REFERENCES

[1] D. Slepian and J.K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Information Theory*, vol. IT-19, no. 4, pp. 471–480, Jul 1973.

[2] A.D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Information Theory*, vol. IT-22, no. 1, pp. 1–10, January 1976.

[3] S.S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," in *IEEE Int. Conf. on Data Compression*, 1999, pp. 158–167.

[4] J. Garcia-Frias, "Compression of correlated binary sources using turbo codes," *IEEE Communications Letters*, vol. 5, no. 10, pp. 417–419, October 2001.

[5] A. Aaron and B. Girod, "Compression with Side Information Using Turbo Codes," in *IEEE Int. Conf. on Data Compression*, April 2002, pp. 252–261.

[6] A.D. Liveris, Z.Xiong, and C.N. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Communications Letters*, vol. 6, no. 10, pp. 440–442, October 2002.

[7] E.H. Adelson and J.R. Bergen, "The plenoptic function and the elements of early vision," in *Computational Models of Visual Processing*. 1991, pp. 3–20, MIT Press.

[8] J.X. Chai, X. Tong, S.C. Chan, and H.Y. Shum, "Plenoptic sampling," in *Computer Graphics (SIGGRAPH'00)*, July 2000, pp. 307–318.

[9] N. Gehrig and P.L. Dragotti, "Distributed compression in camera sensor networks," in *IEEE MMSP-04*, September 2004, Accepted for publication.

[10] N. Gehrig and P.L. Dragotti, "Symmetric and a-symmetric Slepian and Wolf codes with systematic and non-systematic linear codes," Submitted to IEEE Communications Letters, April 2004.